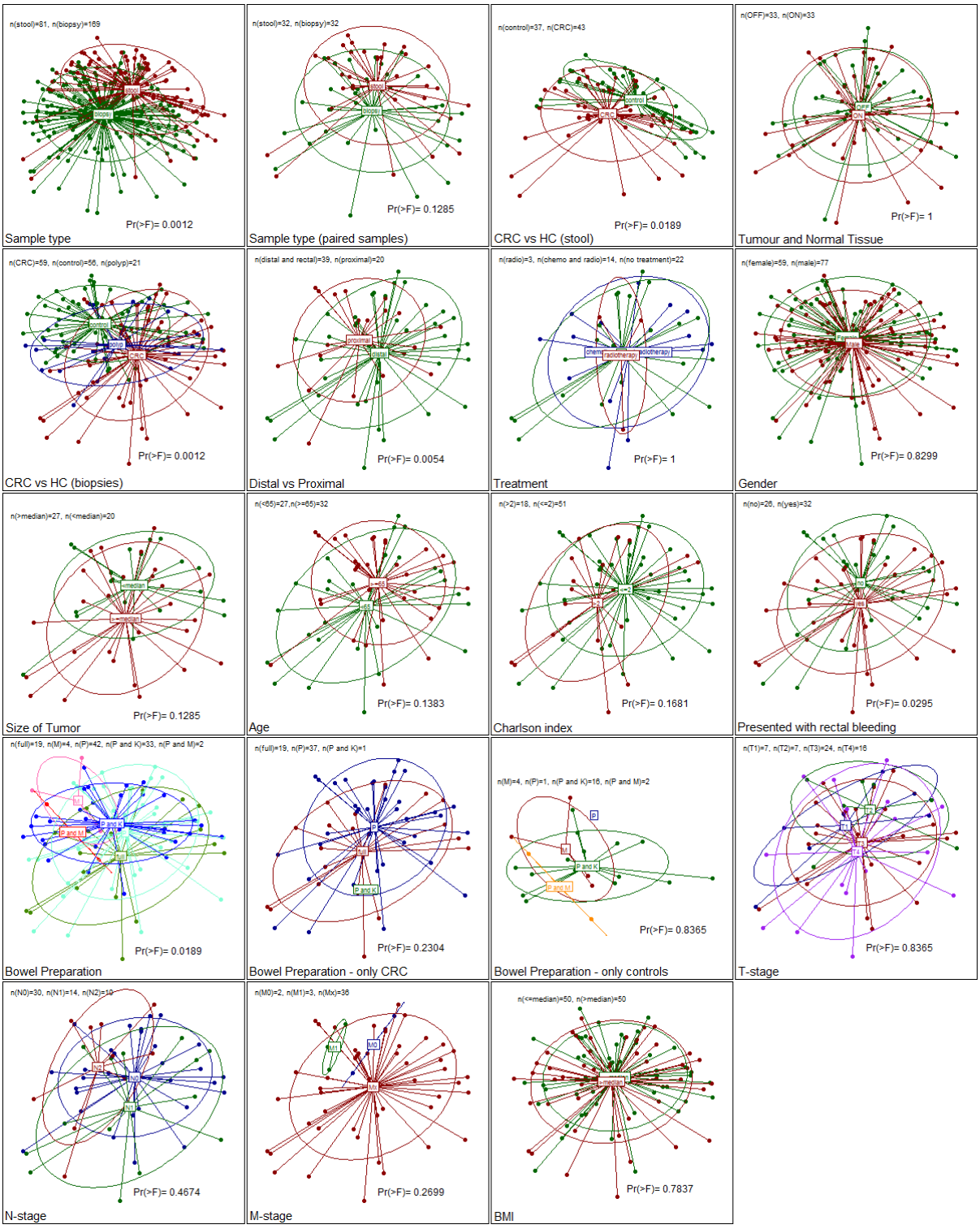
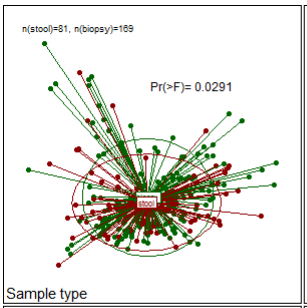
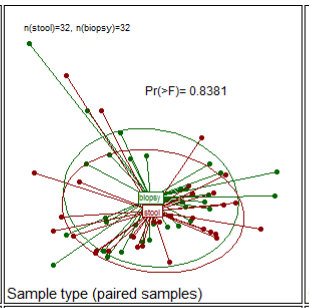
Supplementary Figures 1-4: PCoA of microbiota using different distance matrices (1: Unweighted UniFrac, 2: Weighted UniFrac, 3: Spearman-rank distance, 4: binary distance). Sample groupings according to metadata in Supplementary Table 1.

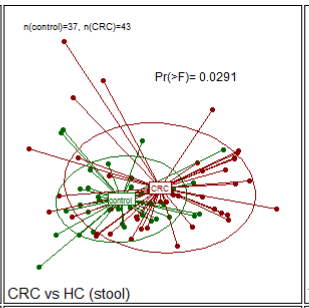




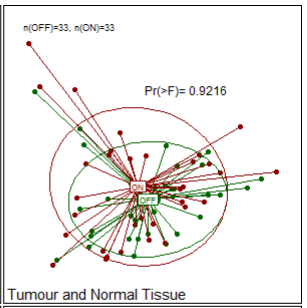
Sample type



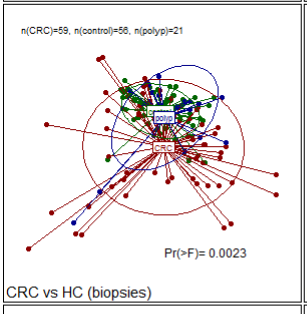
Sample type (paired samples)



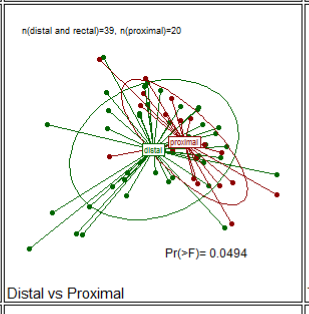
CRC vs HC (stool)



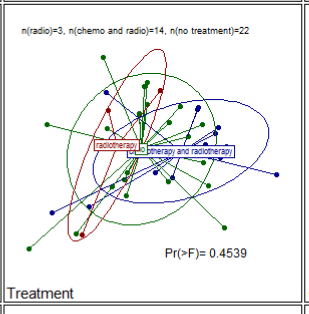
Tumour and Normal Tissue



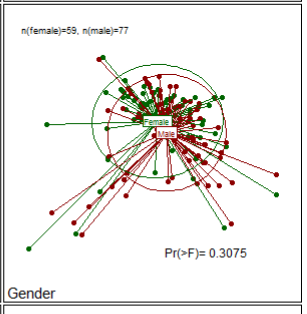
CRC vs HC (biopsies)



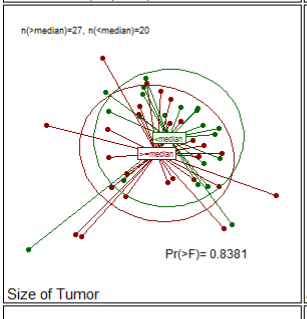
Distal vs Proximal



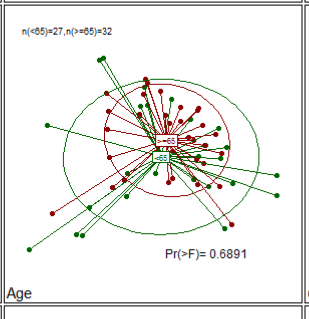
Treatment



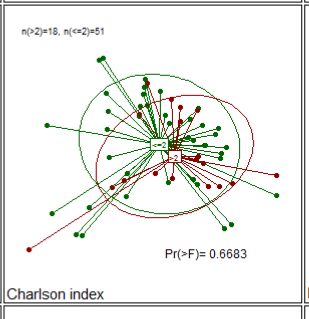
Gender



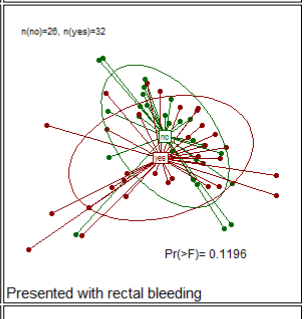
Size of Tumor



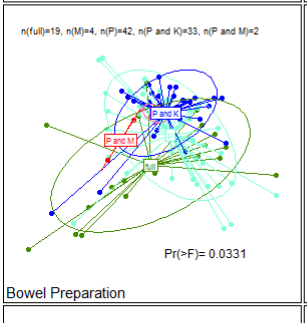
Age



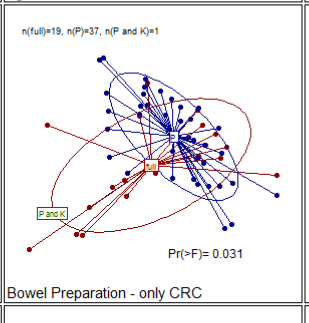
Charlson index



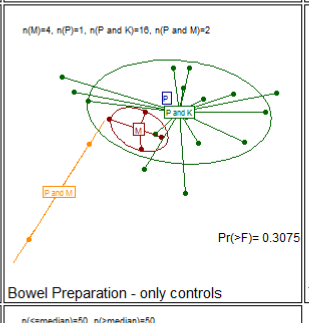
Presented with rectal bleeding



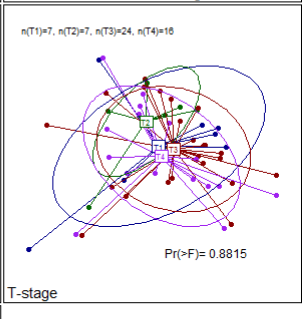
Bowel Preparation



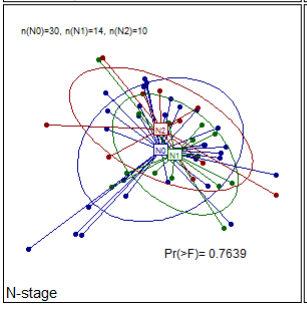
Bowel Preparation - only CRC



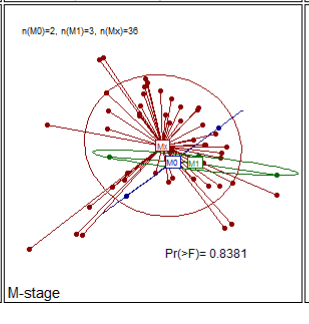
Bowel Preparation - only controls



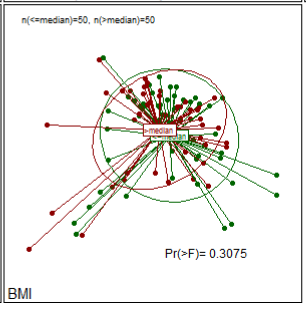
T-stage



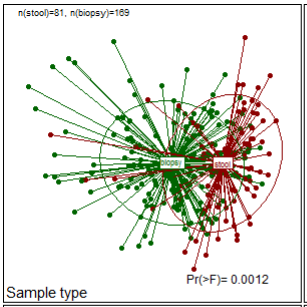
N-stage



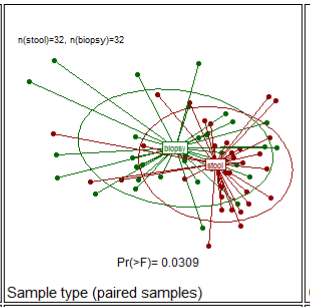
M-stage



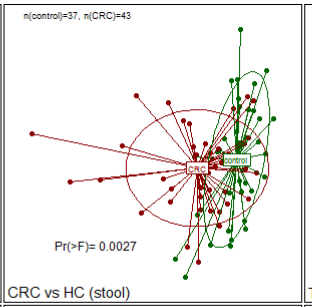
BMI



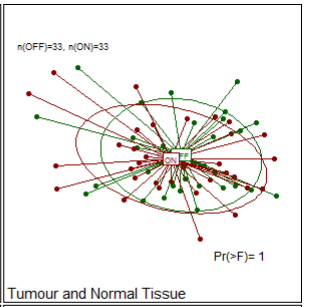
Sample type



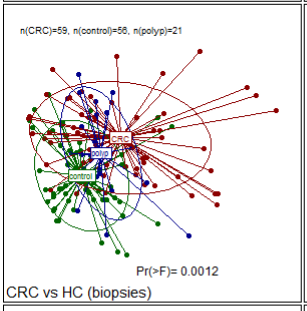
Sample type (paired samples)



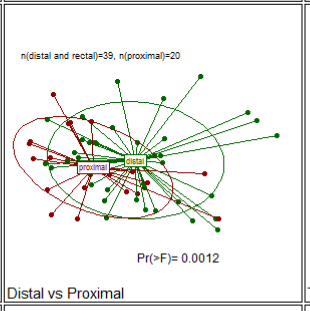
CRC vs HC (stool)



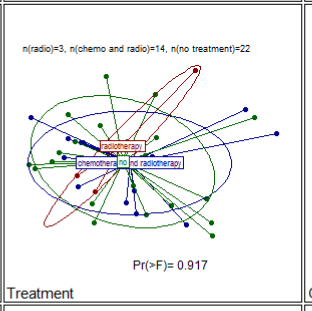
Tumour and Normal Tissue



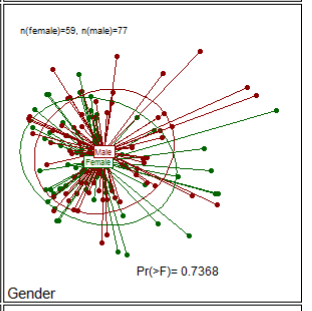
CRC vs HC (biopsies)



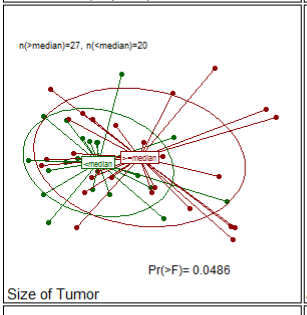
Distal vs Proximal



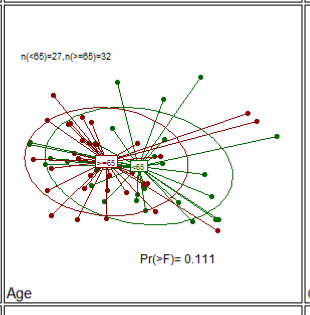
Treatment



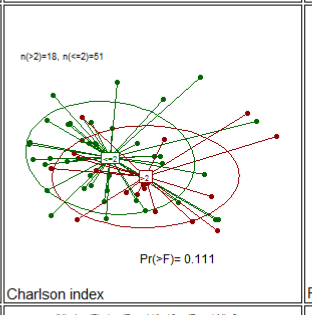
Gender



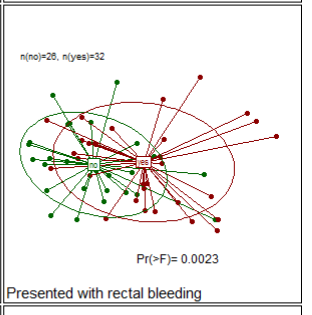
Size of Tumor



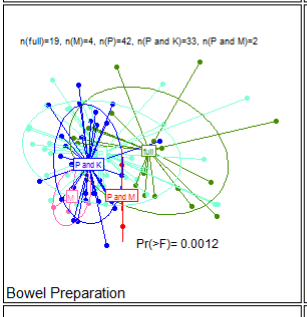
Age



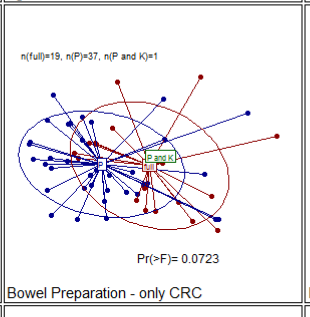
Charlson index



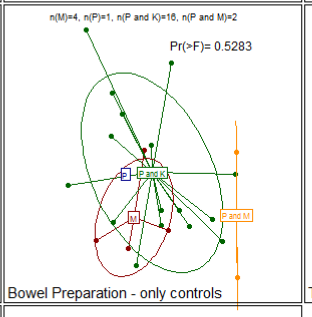
Presented with rectal bleeding



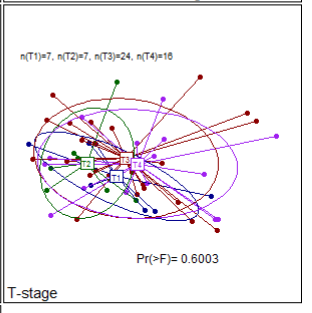
Bowel Preparation



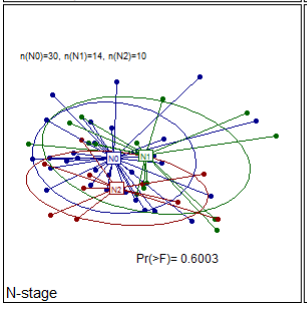
Bowel Preparation - only CRC



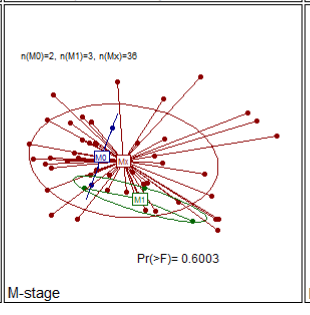
Bowel Preparation - only controls



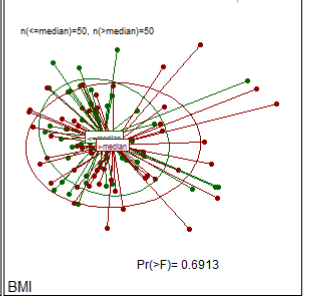
T-stage



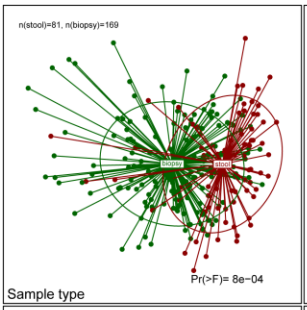
N-stage



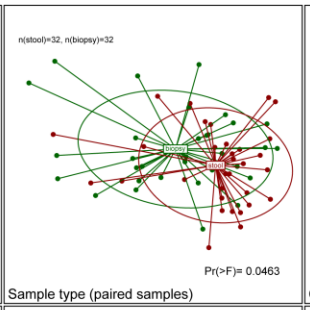
M-stage



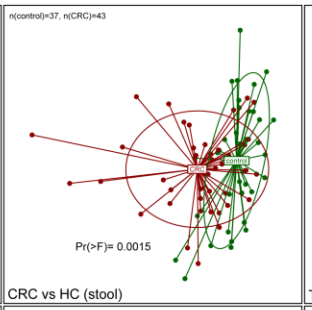
BMI



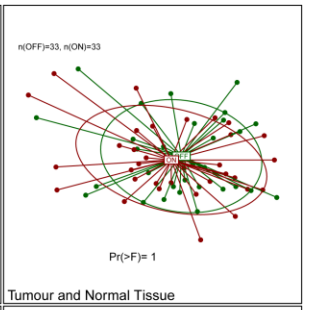
Sample type



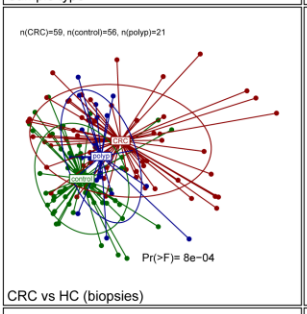
Sample type (paired samples)



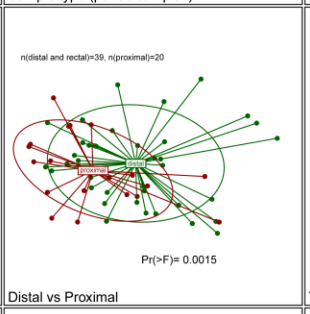
CRC vs HC (stool)



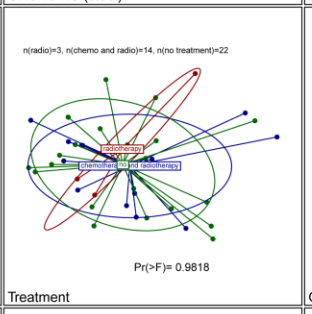
Tumour and Normal Tissue



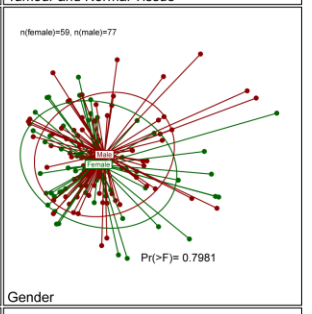
CRC vs HC (biopsies)



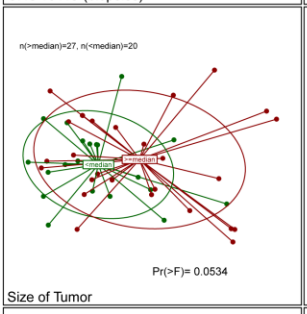
Distal vs Proximal



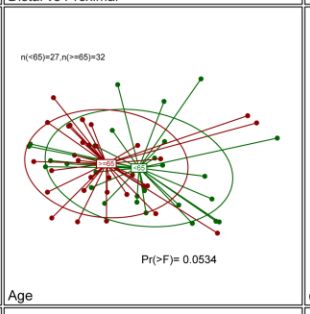
Treatment



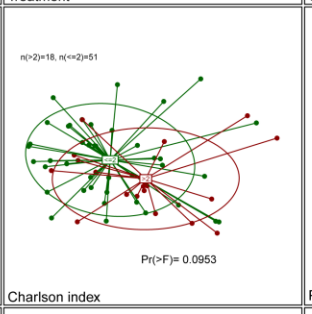
Gender



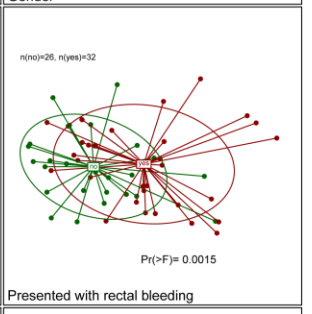
Size of Tumor



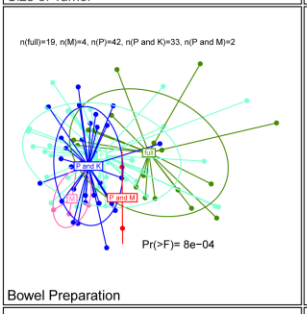
Age



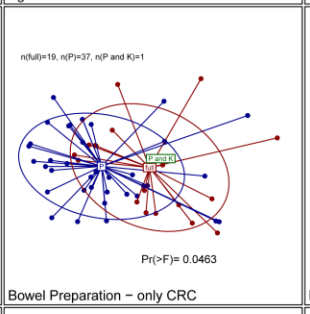
Charlson index



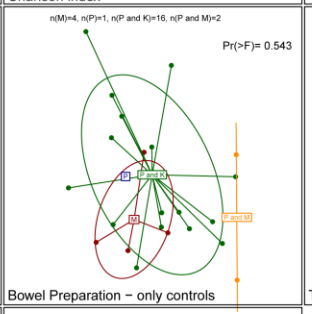
Presented with rectal bleeding



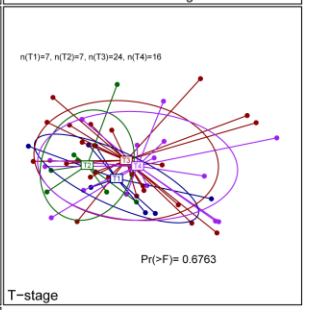
Bowel Preparation



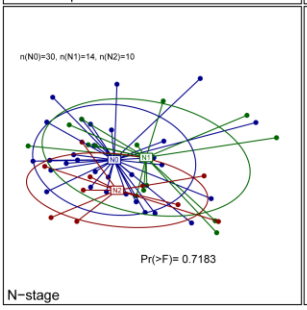
Bowel Preparation − only CRC



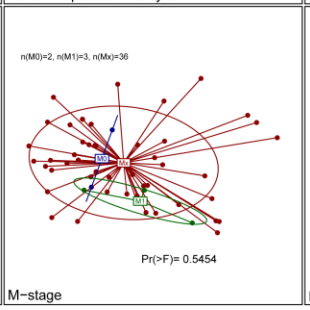
Bowel Preparation − only controls



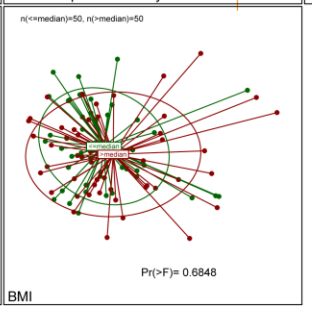
T−stage



N−stage

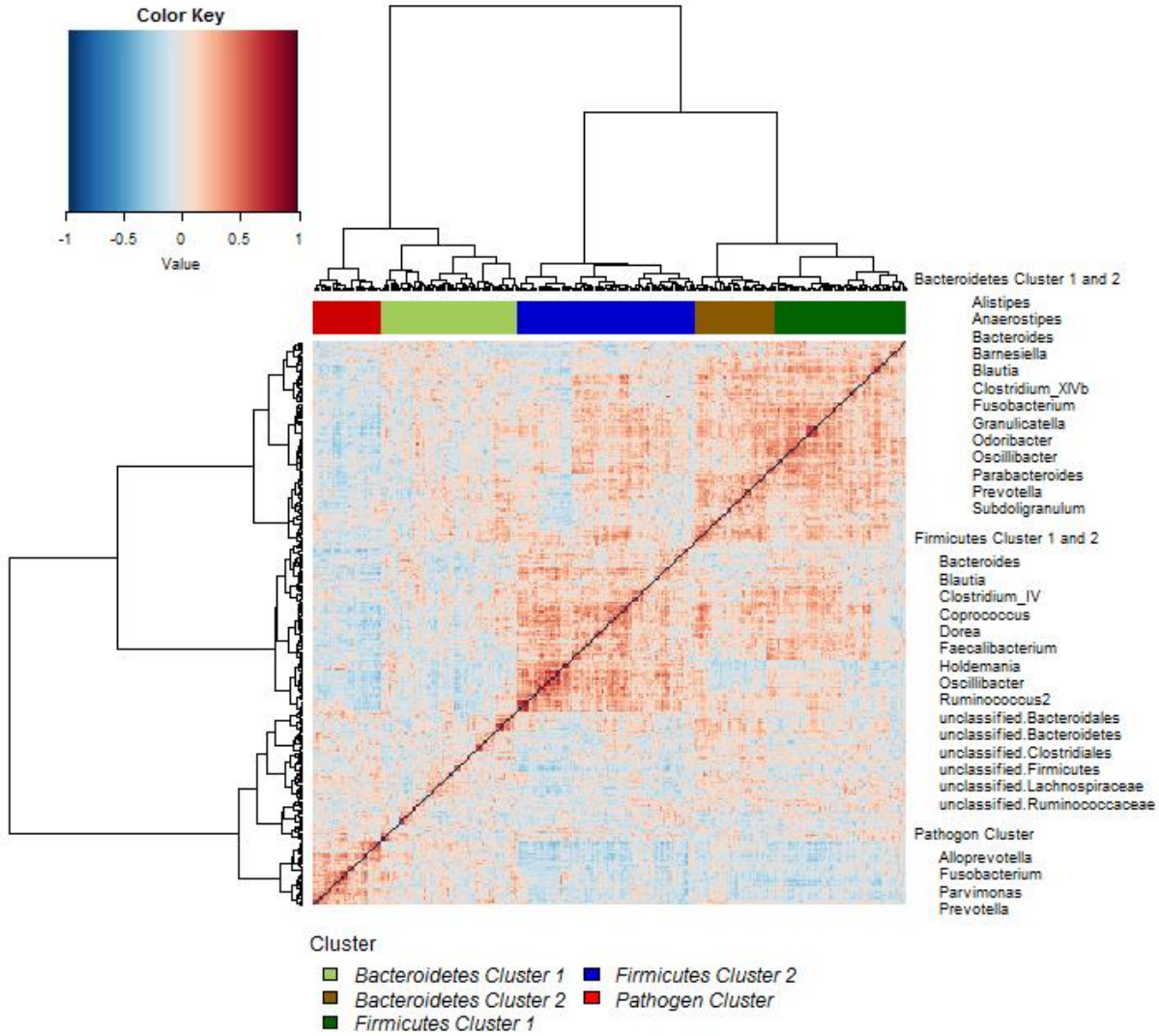


M−stage

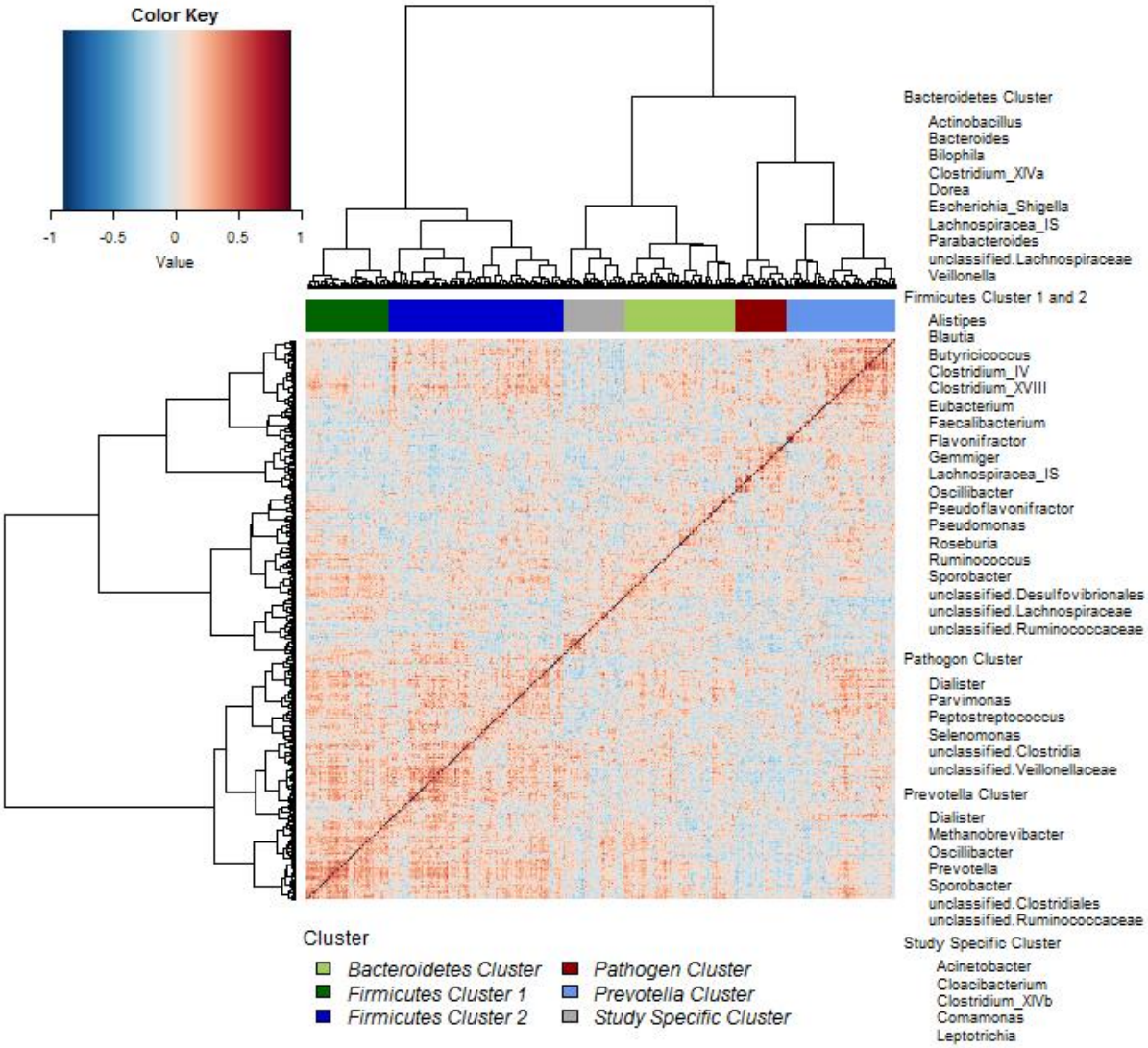


BMI

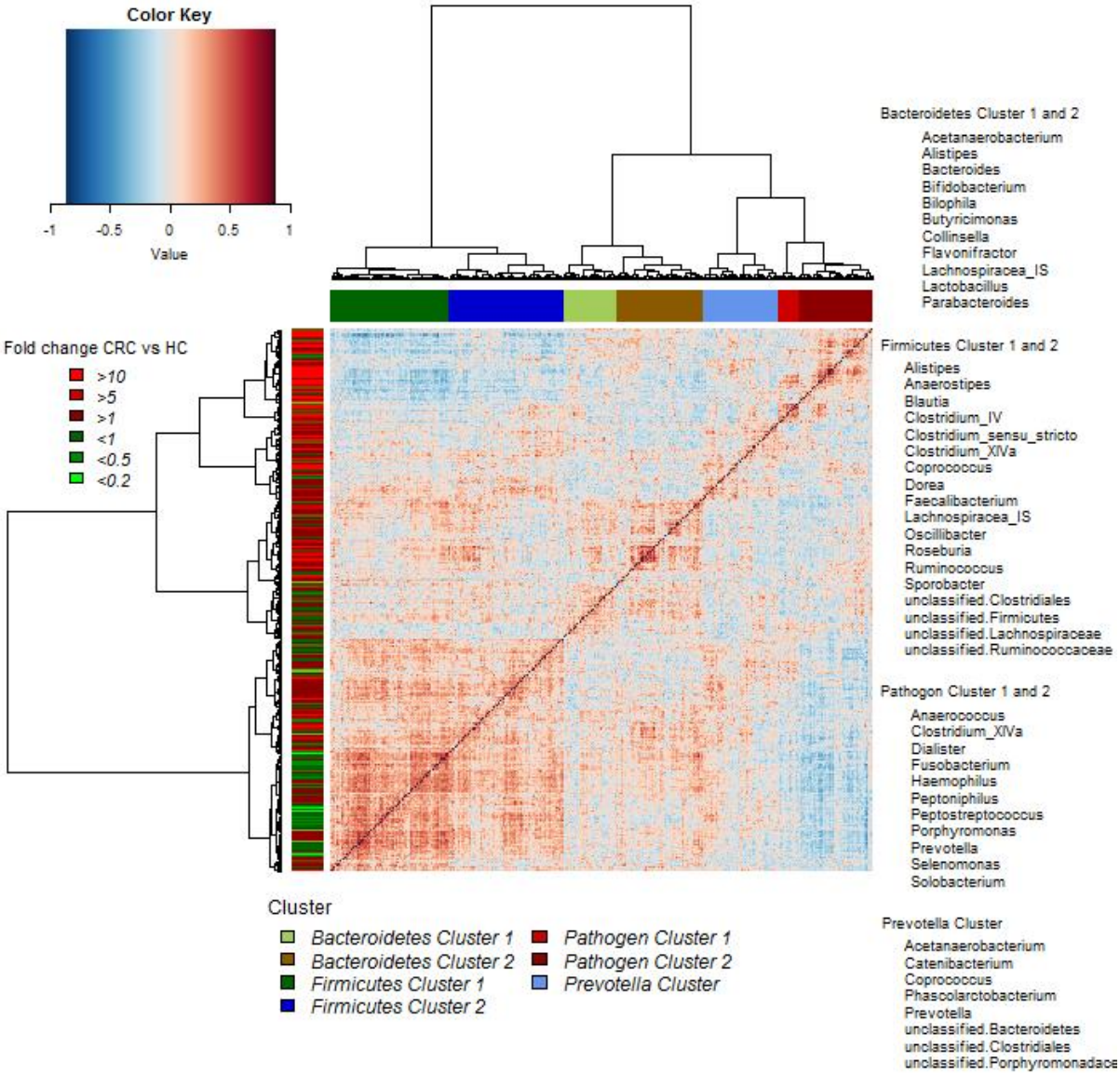
Supplementary Figure 5: Hierarchical Ward-linkage clustering based on the Pearson correlation coefficients of the relative abundance of OTUs in 53 tumour samples from Kostic *et al.*[21]. Co-abundance groups were defined on the basis of the clusters in the vertical tree and named after their most notable characteristic. Column colour coding according to legend below. Row colour coding according to legend on the left. To the right, the most abundant bacterial genera as well as the most strongly connected genera in each CAG (i.e. genera with the highest numbers of significant positive correlations with other members of each respective group) are listed.



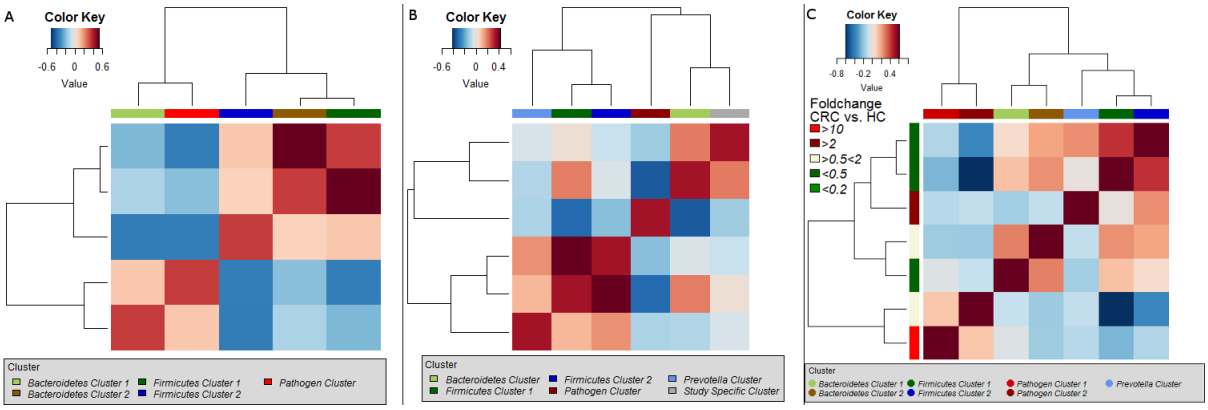
Supplementary Figure 6: Hierarchical Ward-linkage clustering based on the Pearson correlation coefficients of the relative abundance of OTUs in 48 tumour samples from Zeller *et al.* Zeller *et al.*[11]. Co-abundance groups were defined on the basis of the clusters in the vertical tree and named after their most notable characteristic. Column colour coding according to legend below. Row colour coding according to legend on the left. To the right, the most abundant bacterial genera as well as the most strongly connected genera in each CAG (i.e. genera with the highest numbers of significant positive correlations with other members of each respective group) are listed.



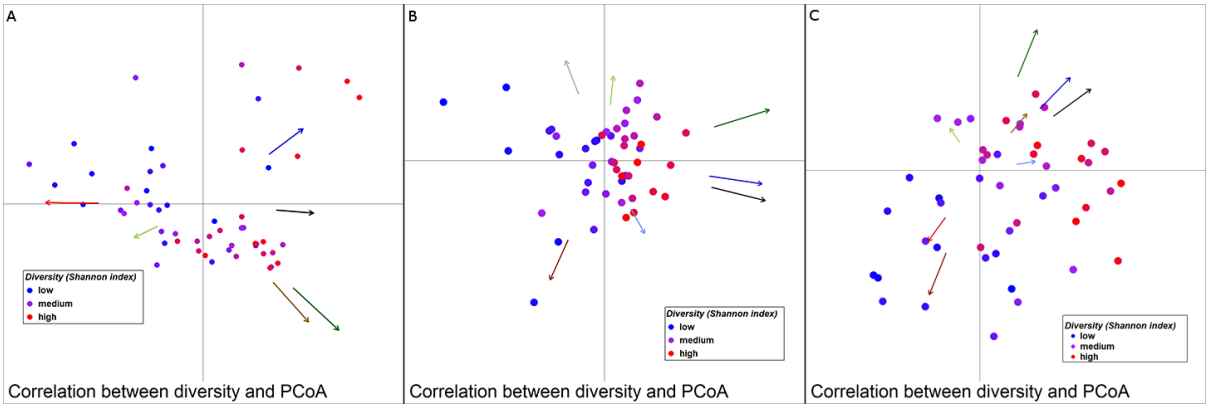
Supplementary Figure 7: Hierarchical Ward-linkage clustering based on the Pearson correlation coefficients of the relative abundance of OTUs in 47 tumour samples from the here presented cohort. Co-abundance groups were defined on the basis of the clusters in the vertical tree and named after their most notable characteristic. Column colour coding according to legend below. Row colour coding according to legend on the left. To the right, the most abundant bacterial genera as well as the most strongly connected genera in each CAG (i.e. genera with the highest numbers of significant positive correlations with other members of each respective group) are listed.



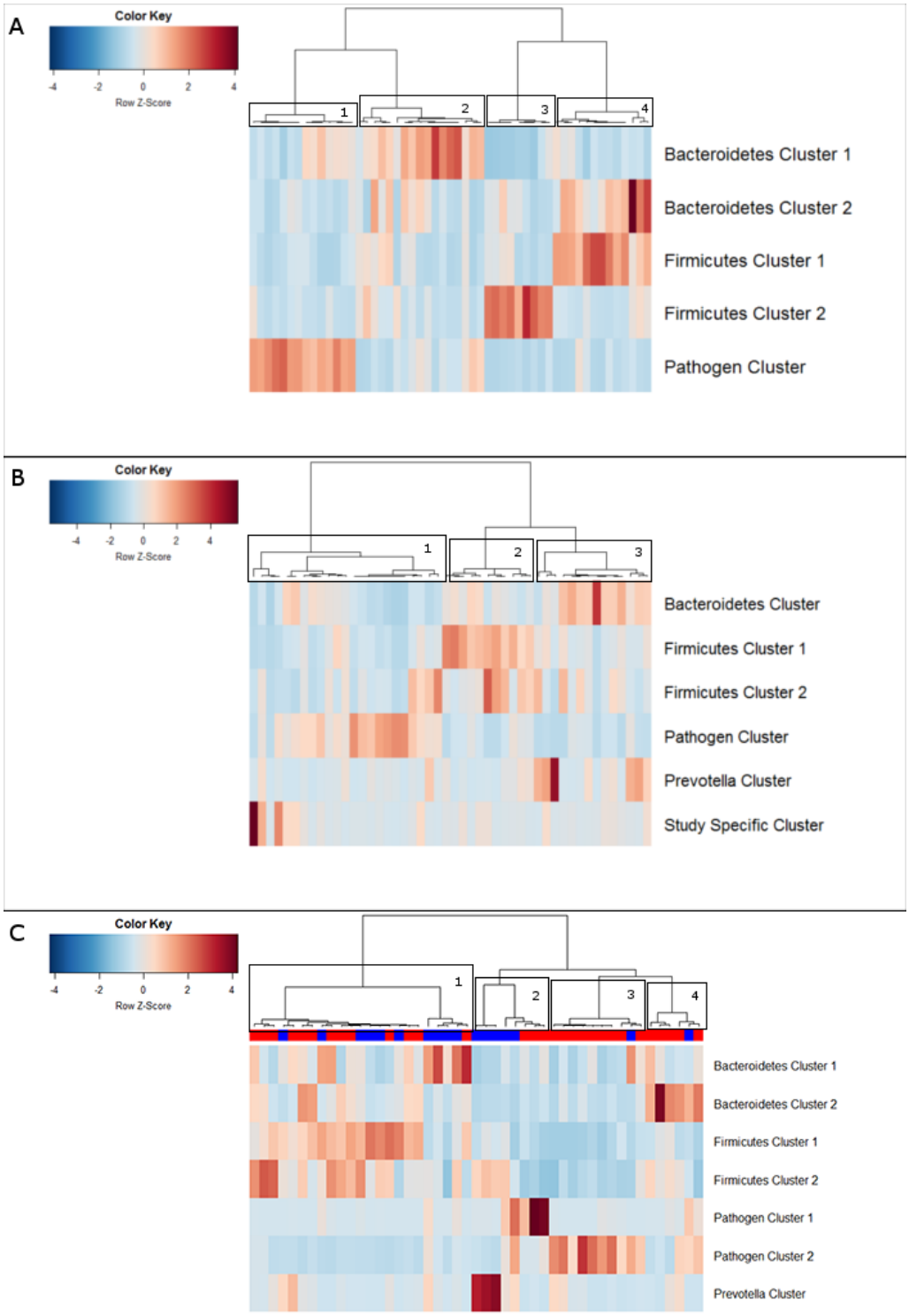
Supplementary Figure 8: Hierarchical Ward-linkage clustering based on the Pearson correlation coefficients of the the relative abundance of CAGs in 53 CRC samples from Kostic *et al.*[21] (A) in 48 CRC samples from Zeller *et al.*[11] (B), and in 47 tumour samples from this study (C).



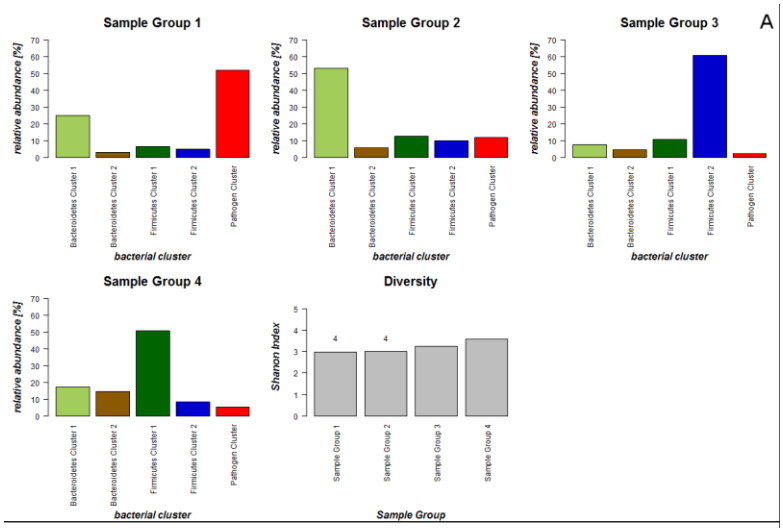
Supplementary Figure 9: Unweighted UniFrac PCoA; A: 53 CRC samples from Kostic *et al.*[21]; B: 48 CRC samples from Zeller *et al.* Zeller *et al.*[11]; C: 47 tumours samples from this study. The location of samples on the PCoA is strongly associated with α-diversity and abundance of the bacterial co-occurrence clusters as defined in Supplementary Figures 5-7. Arrows indicate the direction of correlations for α-diversity (black) and bacterial co-occurrence networks (colours as in Supplementary Figures 5-7) with location on the PCoA. The distance from the origin and the direction correspond to the vector of x- and y-axis correlation.

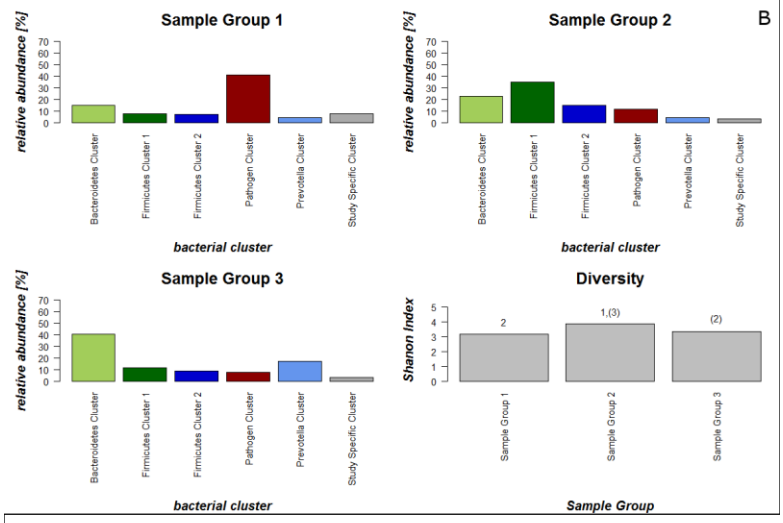


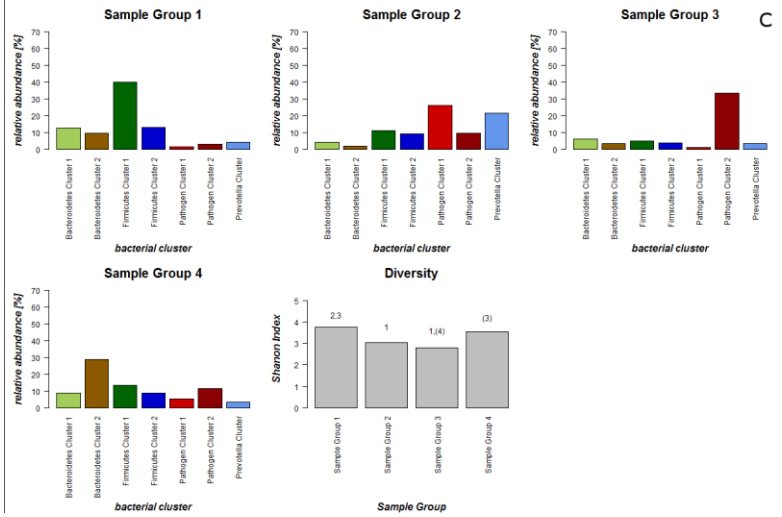
Supplementary Figure 10: Hierarchical Ward-linkage clustering of based on the Pearson correlation of the abundance of bacterial co-occurrence clusters in each sample. (A) 53 CRC samples from Kostic *et al.*[21]; (B) 48 CRC samples from Zeller *et al.*[11]; (C) 47 tumour samples from this study.



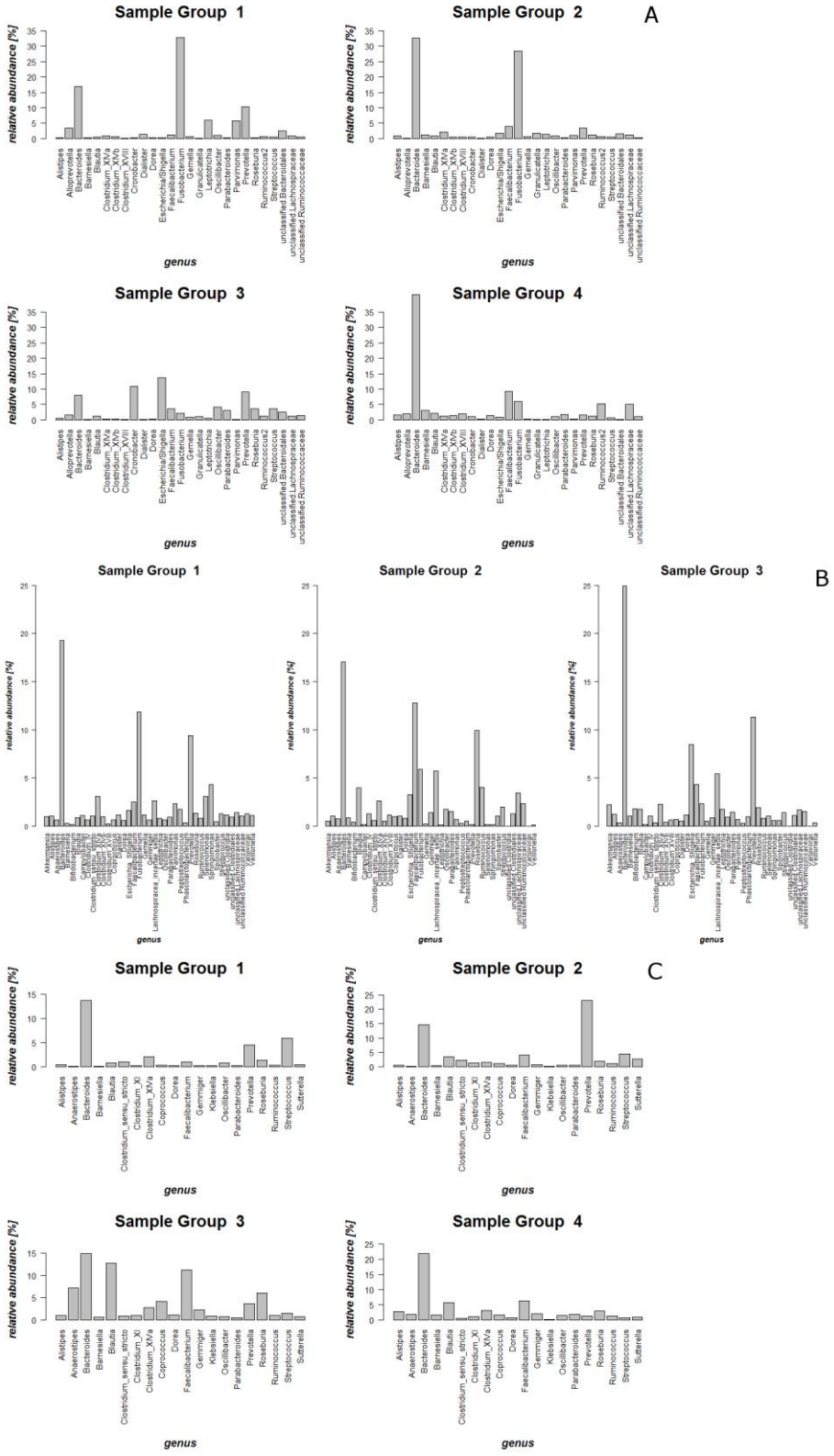
Supplementary Figure 11: Schematic representation of relative abundance distribution for each bacterial co-occurrence cluster; (A) for 53 CRC samples from Kostic *et al.*[21]; (B) for 48 CRC samples from Zeller *et al.*[11]; (C) for 47 tumour samples from this study.Significant difference for each Sample Group compared to each other Sample Group in terms of α-diversity is indicated above bar. Brackets indicate $p<0.1$.


A
Sample Group 1
Sample Group 2
Sample Group 3
Sample Group 4
Diversity
relative abundance [%]
Shanon Index
Bacteroidetes Cluster 1
Bacteroidetes Cluster 2
Firmicutes Cluster 1
Firmicutes Cluster 2
Pathogen Cluster
bacterial cluster
Sample Group 1
Sample Group 2
Sample Group 3
Sample Group 4
Sample Group

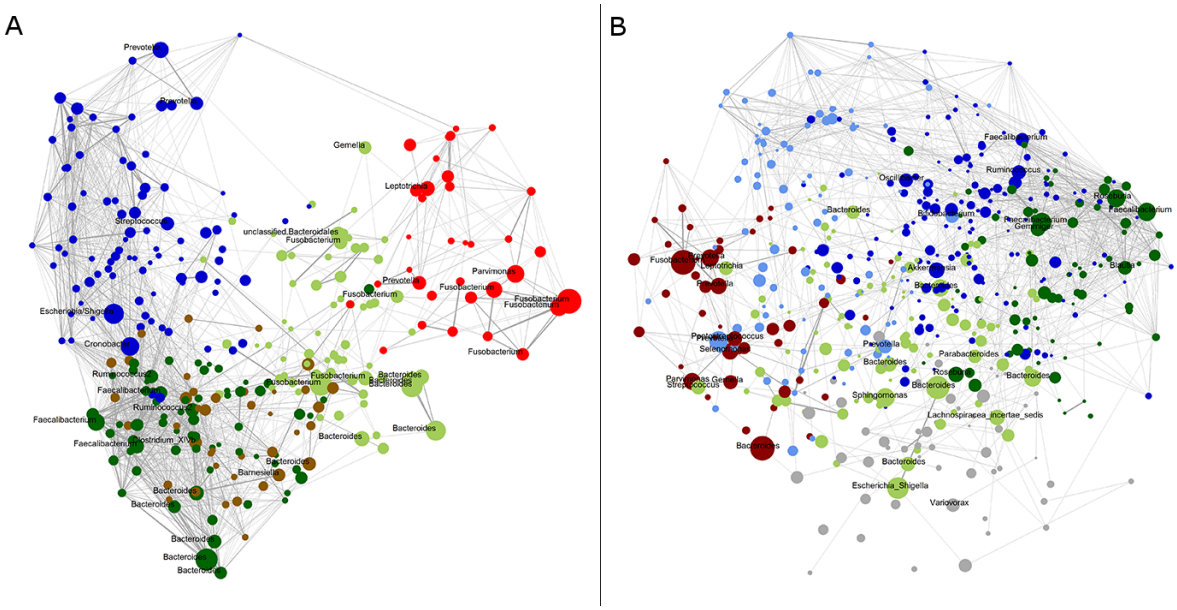

B
Sample Group 1
Sample Group 2
Sample Group 3
Diversity
relative abundance [%]
Shanon Index
Bacteroidetes Cluster
Firmicutes Cluster 1
Firmicutes Cluster 2
Pathogen Cluster
Prevotella Cluster
Study Specific Cluster
bacterial cluster
Sample Group 1
Sample Group 2
Sample Group 3
Sample Group


C
Sample Group 1
Sample Group 2
Sample Group 3
Sample Group 4
Diversity
relative abundance [%]
Shanon Index
Bacteroidetes Cluster 1
Bacteroidetes Cluster 2
Firmicutes Cluster 1
Firmicutes Cluster 2
Pathogen Cluster 1
Pathogen Cluster 2
Prevotella Cluster
bacterial cluster
Sample Group 1
Sample Group 2
Sample Group 3
Sample Group 4
Sample Group

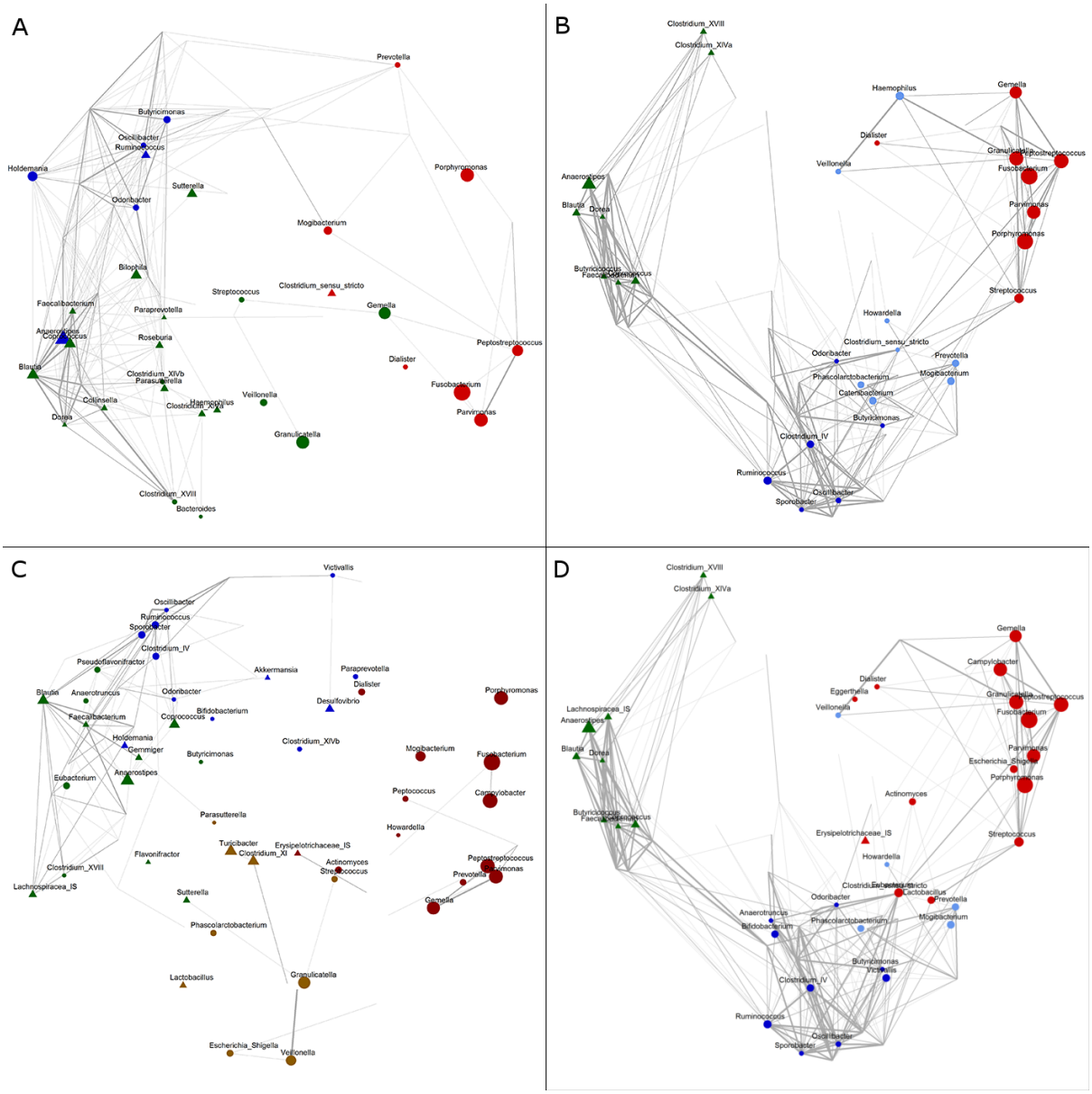
Supplementary Figure 12: Schematic representation of relative abundance distribution for genera with a mean abundance > 0.5%; (A) 53 CRC samples from Kostic *et al.*[21]; (B) 48 CRC samples from Zeller *et al.*[11]; (C) 47 tumour samples from this study.



Supplementary Figure 13: Network plots of OTUs. (A) 53 CRC samples from Kostic *et al.*[21]; (B) 48 CRC samples from Zeller *et al.*[11]. The size of each node (circle) is proportional to the mean abundance of each OTU. The location of each node was determined by a PCoA of the correlation distance as described in Material and Methods. Colour of each node according to the CAGs as in Supplementary Figures 5-7. The width of each edge corresponds to the p-value of the correlation between each respective node (lower p-value, higher line-width); only edges with positive correlations ($p<0.1$) shown. Names only shown for OTUs with a mean abundance > 0.5%.



Supplementary Figure 14: Network plots of genera. (A and B): 53 CRC samples from Kostic *et al.*[21] (A); 47 tumour samples from this study (B). Colour of each node according to the bacterial co-occurrence clusters as in Supplementary Figures 5 and 7. (C and D): 48 CRC samples from Zeller *et al.*[11] (C); 47 tumour samples from this study (D). Colour of each node according to the bacterial co-occurrence clusters as in Supplementary Figures 6 and 7. (A-D): The size of each node (circle) is proportional to the fold-change of each genus between CRC samples from each respective cohort and the healthy controls of this study. Only genera with a fold-change > 1.5 and genera which were shared between the Kostic *et al.* dataset and this study (A and B) and the Zeller *et al.* dataset and this study (C and D), respectively, are shown. The location of each node was determined by a PCoA of the correlation distance as described in Material and Methods. The width of each edge corresponds to the p-value of the correlation between each respective node (lower p-value, higher line-width); only edges with positive correlations ($p<0.1$) shown. Circles: increased in individuals with CRC. Triangles decreased in individuals with CRC.


A
Prevotella
Butyricimonas
Oscillibacter
Ruminococcus
Holdemania
Porphyromonas
Sutterella
Odoribacter
Mogibacterium
Bilophila
Streptococcus
Clostridium_sensu_stricto
Faecalibacterium
Paraprevotella
Gemella
Roseburia
Peptostreptococcus
Dialister
Blautia
Clostridium_XlVb
Parasutterella
Fusobacterium
Collinsella
Veillonella
Haemophilus
Parvimonas
Dorea
Granulicatella
Clostridium_XVIII
Bacteroides
B
Clostridium_XVIII
Clostridium_XlVa
Haemophilus
Gemella
Dialister
Veillonella
Granulicatella
Peptostreptococcus
Fusobacterium
Anaerostipes
Blautia
Dorea
Parvimonas
Porphyromonas
Butyricicoccus
Faecalicatenibacter
Streptococcus
Howardella
Clostridium_sensu_stricto
Odoribacter
Prevotella
Phascolarctobacterium
Mogibacterium
Catenibacterium
Butyricimonas
Clostridium_IV
Ruminococcus
Oscillibacter
Sporobacter
C
Victivallis
Oscillibacter
Ruminococcus
Sporobacter
Clostridium_IV
Pseudoflavonifractor
Akkermansia
Paraprevotella
Dialister
Blautia
Anaerotruncus
Odoribacter
Desulfovibrio
Porphyromonas
Faecalibacterium
Coprococcus
Bifidobacterium
Holdemania
Clostridium_XlVb
Gemmiger
Butyricimonas
Mogibacterium
Fusobacterium
Eubacterium
Anaerostipes
Peptococcus
Campylobacter
Parasutterella
Howardella
Turicibacter
Erysipelotrichaceae_IS
Flavonifractor
Clostridium_XI
Actinomyces
Streptococcus
Peptostreptococcus
Prevotella
Porphyromonas
Clostridium_XVIII
Lachnospiracea_IS
Sutterella
Gemella
Phascolarctobacterium
Lactobacillus
Granulicatella
Escherichia_Shigella
Veillonella
D
Clostridium_XVIII
Clostridium_XlVa
Gemella
Campylobacter
Dialister
Eggerthella
Granulicatella
Peptostreptococcus
Veillonella
Fusobacterium
Lachnospiracea_IS
Anaerostipes
Blautia
Dorea
Parvimonas
Escherichia_Shigella
Porphyromonas
Actinomyces
Butyricicoccus
Faecalicatenibacter
Erysipelotrichaceae_IS
Streptococcus
Howardella
Clostridium_sensu_stricto
Lactobacillus
Odoribacter
Prevotella
Anaerotruncus
Bifidobacterium
Phascolarctobacterium
Mogibacterium
Butyricimonas
Victivallis
Clostridium_IV
Ruminococcus
Oscillibacter
Sporobacter

Supplementary Figure 15: PCA of dietary information. The habitual diet of individuals with CRC and polyps was significantly different from the habitual diet of healthy individuals.